

Mining Temporal Association Rules from Time Series Microarray Using Apriori Algorithm

Bhuvaneswari V.¹, Umajothy P.²

Assistant Professor¹, M Phil Research Scholar²

Department of Computer Applications, School of Computer Science and Engineering, Bharathiar University, Coimbatore, Tamil Nadu, India

bhuvanes_v@yahoo.com; umajo7@gmail.com

Abstract

Microarray Technology is used to automate the diagnostic task and improve the accuracy of the traditional diagnostic techniques. Analysis of microarrays presents a number of unique challenges for data mining. It is a high throughput method to analyze expression levels of multiple genes simultaneously. The concept of association plays an important role in microarray to find the association of expression levels of multiple genes. In the proposed work, the association rule mining technique is used to mine gene expression data in order to analyze the affects of expression of one gene with other gene for gene functionality biological Process and Molecular Function. The yeast *Saccharomyces cerevisiae* dataset is used for the work where Apriori algorithm is used to find the association of genes for different experimental condition for time interval 2 and 4.

Keywords

Association; Discrete; Microarray; Regulations; Support; Temporal; Time Interval; Transaction

Introduction

Data mining is the extraction of hidden predictive information from large databases. Data mining is a larger process known as Knowledge Discovery in databases (KDD). The process of discovering meaningful, new correlation patterns and trends by shifting through large amount of data stored in repositories, using pattern recognition techniques as well as statistical and mathematical techniques. Bioinformatics is the science of managing, mining and interpreting information from biological sequences and structures [1].

Bioinformatics and data mining provide exciting and challenging researches in several application areas especially in computer science. A microarray is a

sequence of dots of DNA, protein, or tissue arranged on an array for easy simultaneous analysis. The DNA microarray plays an integral role in gene expression profiling. Alternative names for DNA microarray are *gene chip*, *DNA chip* and *Biochip* [3]. Gene is a section of DNA at a specific position on a particular chromosome that specifies the amino acid sequence for a protein.

A gene of an organism is composed of many genetic interacting elements. It is a tough process to discover the genetic interacting elements in the complex biological regulations. The microarray technique allows researchers to observe the expression levels of thousands of genes in a single experiment. Gene expression analysis in a microarray experiment is used to monitor the expression levels of genes at a genome scale. Various association algorithms are used to identify the expression levels of thousands of genes simultaneously under a particular condition. The microarray consists of gene values which are interrelated with one another as well as time constraints where time expressions are included. The problem of handling time series data by the proposed method using association rule mining helps to extract the temporal dependency among genes.

An association rule mining is used to find the interesting association and correlation relationships within items in a large database. An association rule is a pair of disjoint itemsets and if LHS and RHS denote the two disjoint itemsets, then it can be written as $LHS \rightarrow RHS$. Support and confidence are the two parameters of the association rule.

- Support measure is to find all the frequent item sets that satisfy minimum support threshold value. This can be represented as

$$\text{Support} = T(A \cup B)$$

- Confidence is used to find all the high confidence rules from frequent item sets in the support values. Confidence of the rule with

$$\text{Confidence} = T(B/A)$$

LHS \rightarrow RHS with the transaction set T is the ratio support (LHS \rightarrow RHS)/ support (LHS).

A temporal association rule represents the various transcriptional time delays between associated genes. It has the form [gene A \uparrow , gene B \downarrow] \rightarrow (7 min) [gene C \uparrow], which represents the high expression level of gene A and B followed by expression of gene C after some minutes [6].

Definition 1: A temporal item is an item which has a time stamp and a temporal item set is a non-empty set of temporal items. A temporal association rule expresses that a set of items tends to appear along with another set of items in the same transactions, in a specific time frame.

Definition 2: A temporal association rule is a pair of disjoint temporal item sets. LHS and RHS denote the left and right temporal item sets. The temporal association rule is written as LHS $> (\Delta)$ RHS, where Δ is the interval of two different time stamps. The temporal association rule extracts the size of the transcriptional time delay of the associated genes (minutes and the activation and inhibition relationship [gene A $\uparrow \rightarrow$ gene C \downarrow] of the co regulation of genes (gene A \uparrow , gene B \downarrow).

In this work, a method is proposed to extract the temporal association rules from the yeast microarray data using Apriori algorithm and analyze the similar pattern for genes in different time series of the experimental data.

The paper is organized as follows. Section 2 describes the literature review of various association rule mining algorithms used to analyze the microarray data. Section 3 describes the proposed methodology for mining the temporal association rules using Apriori algorithm. Section 4 depicts the experimental results of the proposed method. In section 5, conclusion of the work is made.

Related Work

Microarray data presents new challenges which make many traditional data mining techniques infeasible to

extract the hidden gene relationships. The main challenge is its high density –a number of attribute (columns) and a considerably smaller number of expression experiments (rows). To use current data mining algorithms, biologists manage to simplify the complexity of their data by means of building restriction on the analysis to small proportion of attributes. The microarray data consists of genes names with expressions which are interrelated with one another. The association relationship between the genes can be found by association rule mining whose algorithms like CLOSET, Apriori, and Partition algorithm are used to find the frequent item sets and association rules.

Association rule mining (ARM) widely used for mining large databases was originally introduced to handle market basket data for consumer purchasing patterns in various application areas. The exploration of mining microarray data has led to many proposals of mining the association rules.

In [17], an introduction has been made to a data mining technology “Data Mining Ready” with data structure called Peano count tree for Association Rule Mining. Algorithms like, Peano-ARM (Association rule mining) and P-gen algorithm are also used for mining. The microarray data is organized into a bit-sequential format where each bit file is converted to quadrant base P-trees that are easy to derive rules from microarray data.

In [1], the problem of mining association rules has been discussed. The comparison of the algorithms like Apriori and Apriori TID and a new algorithm called Apriori hybrid have been proposed. The efficiency of the algorithms is tested and compared with that of the algorithms AIS (Artificial Immune system) for mining large databases. In [14], an introduction on a study on the analysis of DNA microarray data has been depicted using the association rules. Association rule mining algorithms like Apriori, FP tree growth, Partition, Dynamic-FP growth, Dynamic Item set counting (DIC) are to analyze and associate the gene expression data. The algorithms are explained to find the frequent item sets and association rules. In [15], an APD method called the Association Pattern Discovery has been put forward to discover frequent item sets and association rules. The co-regulated gene profiles are discovered by MAP (Mining attribute Profile) algorithm in the yeast microarray datasets that is compared with the traditional APD methods and resulted in the fact that the MAP was proved to be the

best performance. In [16], a new approach for mining FIS (Frequent item set) tree mining algorithm was introduced making use of bit string data partition format to propose association relationships among different genes by means of two data structures called BSC and FIS tree, the first of which acts as a string compression tree for each bit string representing each gene while FIS tree is to store all the frequent item sets.

In [2], an incremental mining of association rules was introduced using an extended TFP Apriori tree (Total from Partial) that in exploration handles the problem of mining association rules in incremental databases by building TFP trees incrementally. In [6], a temporal association rule mining method has been proposed to extract the temporal dependencies among the related genes using the Apriori algorithm. The inferences of the results are done by Gene Ontology and KEGG pathway.

In [8], a new algorithm called T-Apriori has been put forward which extracts the temporal association rules with respect to time. The frequent item sets are extracted with its support and confidence values. In [12], an algorithm called SPFA (Standing for Segmented Progressive Filter Algorithm) was introduced for large databases, generating the temporal frequent item sets, and temporal sub items. The first part of the algorithm divides the database into partitions, and the second part filters the 2-item sets. The algorithm minimizes the execution time by using scan reduction technique generating all the candidate item sets. In [18], a regularized neural network model was described for characterization of the multiple heterogeneous temporal dynamic patterns of gene expression. A feed forward neural network model is introduced to model the gene expressions. The method is performed in yeast microarray data and compared with the Nearest Neighbor, SVM (support vector machine), and self organized maps resulting in a best performance method.

Methodology

The objective of the proposed work is to extract the temporal association rules for yeast microarray dataset from different experimental Conditions using Apriori algorithm. The framework for the work is given in FIG 1.

The Framework consists of three phases. In the first phase, the preprocessing of microarray data is done, in the second phase, the genes are classified based on the functionality using GO ontology. The temporal

association rules are extracted for the intervals 2 and 4 and the similar pattern of rules are extracted and compared.

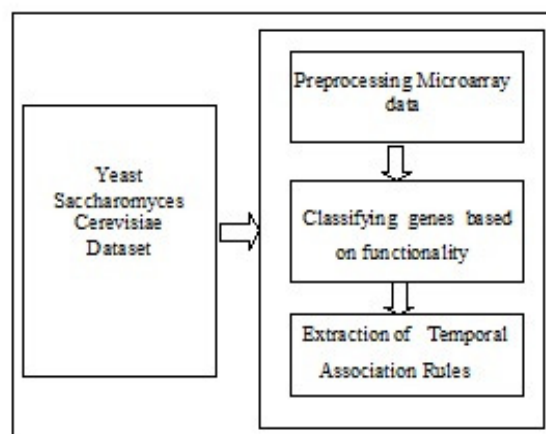


FIG. 1 EXTRACTION OF TEMPORAL ASSOCIATION RULES

Preprocessing Microarray Data

The yeast microarray data is used as dataset to implement the proposed work. The Saccharomyces cerevisiae yeast data consists of 6400 genes in which empty spots and null values are inclusive. In the preprocessing step, the genes with empty spots and null values are removed using the Knimpute method. After that the dataset has 6314 genes which are used to analyze the temporal patterns.

Classifying Genes Based on Functionality

The yeast microarray data is classified based on the functionality using Gene Ontology (GO) which is a controlled vocabulary used to describe the function of gene Products and divided into three functionalities: Molecular Function (MF), Biological Process (BP) and Cellular Component (CC). Molecular Function is used to describe the function of gene product at molecular level; Biological Process describes the participation of gene in biological activities. Cellular component depicts the location of the gene product at cell level. The gene in microarray data is mapped to the gene functionality using the SGDann structure.

SGDann, a master structure of Yeast microarray data contains the parameters namely SGD aspect, SGDgenes, SGDgo. The entire microarray data is mapped with this structure (SGDgenes) and the gene expression in the microarray data is partitioned based on the functionality of gene for Biological Process and Molecular Function with corresponding time series values. 1079 gene expression data for Biological Process and 1153 for Molecular Function extracted are used to find the temporal association of genes.

Association of Microarray Genes

Association rule mining is the one that discovers the frequent patterns, associations and co-relations of item sets which are meaningful to the users and generates the strong rules on the basis of frequent patterns. Association rule mining is used to identify the expression of one particular gene affecting the expression of other genes. The gene expression data for Molecular Function and Biological Process employed to find the gene association for time interval 2 and 4 consists of two phases, first of which is the conversion of discrete values and the other is the extraction of temporal associations among genes.

(1) Conversion of Gene Values to Discrete Values

The gene expression values is converted to discrete values as up regulated and down regulated gene to extract the temporal association rules for the time intervals 2 and 4. Up-regulation is the increase in expression of a gene in which the transcription of a specific mRNA is increased. It is denoted by the symbol "U". Down-regulation is the decrease in the number of receptors for a Chemical or drug on cell surfaces in a given area; and it is denoted by the symbol "D". The original gene expression matrix is converted to discrete values as up regulation and down regulation using the Eq (1) and Eq (2) shown in Table 1 and Table 2.

If (Gene value >0) then "U" ----- Eq (1)
If (Gene value <0) then "D" ----- Eq (2)

TABLE 1 GENE EXPRESSION VALUES

Microarray genes / Time Interval	T0	T1	T2	T3	T4	T5	T6
YARO29W	-0.4700	0.2440	-0.3990	0.1370	-0.3110	0.3430	0.0910
YARO62W	-0.3820	0.2360	0.1180	0.0980	-0.4400	0.3120	0.9300
YARO68W	-0.1370	0.4050	0.5410	0.6350	0.3080	0.5970	1.0340

TABLE 2 DISCRETE VALUES

Microarray genes/Time Interval	T0	T1	T2	T3	T4	T5	T6
YARO29W	D	U	D	U	D	U	U
YARO62W	D	U	U	U	D	U	U
YARO68W	D	U	U	U	U	U	U

(2) Extracting Temporal Associations

The discretized data is used as input to extract temporal association of gene for time interval 2 and 4 for Molecular Function and Biological Process.

The time series value of each gene ranges from T0, T1, T2, T3, T4, T5, and T6. The time stamps of the transactions sets for the Molecular Function and Biological Process for the time interval $\Delta=2$ are set as T0+T2, T1+T3, T2+T4, T3+T5, T4+T6 and time interval $\Delta=4$ is set as T0+T1+T2+T3, T1+T2+T3+T4, T2+T3+T4+T5, T3+T4+T5+T6. The association rule mining Apriori algorithm is used to extract the temporal association for the time stamps in interval 2 and 4. The gene expression data is analyzed for various support and confidence measures. The rules are extracted and the association of gene is analyzed to identify the similar co-occurrence of genes in Molecular Function and Biological Process for time intervals 2 and 4. The pseudo code of the algorithm is shown below.

Pseudo Code

```

Ck= candidate item set of size K
Lk= frequent item set of size K
L1= Frequent item sets;
For (K=1; Lk=; k++)
Ck+1= item set generated from Lk;
For all transactions t, do
Increment the count Ck+1 in t;
Lk+1= candidates in Ck+1 with min_support
End
Return U Lk Lk:

```

The transaction sets is extracted with its gene name with corresponding discrete values. The value consists of up and down regulated genes at different time intervals. The first gene in the Table 2 is taken and associated with other forth coming genes in the same interval.

TABLE 3 TEMPORAL ASSOCIATION OF GENES FOR TIME INTERVAL

Interval	Molecular Function	Interval	Biological Process
T0-T2 YARO29W(D) YARO62W(U) YARO68W(D)	YARO62W(D), YARO62W(U), YARO68W(D), YARO68W(U) YARO68W(D), YARO68W(U)	T0-T2 YARO29W(D) YARO62W(D) YARO68W(D)	YARO62W(D), YARO62W(U), YARO68W(D), YARO68W(U) YARO68W(D), YARO68W(U)
T1-T3 YARO29W(U) YARO62W(U) YARO68W(U)	YARO62W(U), YARO62W(U), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)	T1-T3 YARO29W(U) YARO62W(U) YARO68W(U)	YARO62W(U), YARO62W(U), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)
T2-T4 YARO29W(D) YARO62W(U) YARO68W(U)	YARO62W(U), YARO62W(D), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)	T2-T4 YARO29W(D) YARO62W(U) YARO68W(U)	YARO62W(U), YARO62W(D), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)
T3-T5 YARO29W(U) YARO62W(U) YARO68W(U)	YARO62W(U), YARO62W(U), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)	T3-T5 YARO29W(U) YARO62W(U) YARO68W(U)	YARO62W(U), YARO62W(U), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)
T4-T6 YARO29W(D) YARO62W(D) YARO68W(D)	YARO62W(D), YARO62W(U), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)	T4-T6 YARO29W(D) YARO62W(D) YARO68W(D)	YARO62W(D), YARO62W(U), YARO68W(U), YARO68W(U) YARO68W(U), YARO68W(U)

Likewise in each transaction, the gene name with its values in the intervals i.e., T1+T3, T2+T4... T4+T6 are checked. A relation which occurs in the intervals for support 50, 80,100 is taken as the frequent item sets for the time interval 2 and 4. The below Table 3 and Table 4 shows the sample of 3 genes with the association for the intervals $\Delta=2$ and $\Delta=4$ for Molecular function and Biological Process.

TABLE 4 TEMPORAL ASSOCIATION OF GENES FOR TIME INTERVAL 4

Interval	Molecular Function	Interval	Biological Process
T0+T1-T2+T3 YARO2FW(D)	YARO6ZW(D), YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)	T0+T1-T2+T3 YARO2FW(D)	YARO6ZW(D), YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)
YARO6ZW(D)	YARO6SW(D), YARO6SW(U), YARO6SW(U), YARO6SW(D)	YARO6ZW(D)	YARO6SW(D), YARO6SW(U), YARO6SW(U), YARO6SW(D)
T1+T2-T3+T4 YARO2FW(U)	YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6ZW(D), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)	T1+T2-T3+T4 YARO2FW(U)	YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6ZW(D), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)
YARO6ZW(U)	YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)	YARO6ZW(U)	YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)
T2+T3-T4+T5 YARO2FW(D)	YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)	T2+T3-T4+T5 YARO2FW(D)	YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6ZW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)
YARO6ZW(U)	YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)	YARO6ZW(U)	YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)
T3+T4-T5+T6 YARO2FW(U)	YARO6ZW(U), YARO6ZW(D), YARO6ZW(U), YARO6ZW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U), YARO6SW(U)	T3+T4-T5+T6 YARO2FW(U)	YARO6ZW(U), YARO6ZW(D), YARO6ZW(U), YARO6ZW(U), YARO6SW(U), YARO6SW(D), YARO6SW(U), YARO6SW(U)
YARO6ZW(U)	YARO6SW(U), YARO6SW(U), YARO6SW(D), YARO6SW(D), YARO6SW(U), YARO6SW(U)	YARO6ZW(U)	YARO6SW(U), YARO6SW(U), YARO6SW(D), YARO6SW(D), YARO6SW(U), YARO6SW(U)

The proposed method extracts the temporal association rules from the temporal Patterns for the Molecular Function and Biological Process for the time intervals 2 and 4 which are explained in experimental results.

Experimental Results

Dataset

The results of the proposed work are analyzed for temporal association among genes in the dataset from Molecular Function and Biological Process and then compared to find the similar Pattern of rules.

The Yeast *Saccharomyces Cerevisiae* Dataset is downloaded from National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) website for the proposed work. The Yeast microarray data contains about 6400 Genes (e.g. YALO51W, YALO54C) with their corresponding Yeast values (0.1650, 0.2720). The file “yeastgenes.sgd” is obtained from the GO annotation site. The number of genes in the dataset is shown in the below Table 5.

TABLE 5 DATASET

Dataset	Number of Genes
SGD genes	37794
Yeast microarray	6400
After removal of empty genes	6413

The yeast microarray dataset contain 6400 genes. After

removal of empty genes from the dataset, there are about 6314 genes which are shown in FIG 2. The entire microarray dataset is mapped based on the functionality. After mapping, 1079 has been acquired for Biological process, 1153 for Molecular function is shown in FIG 3.

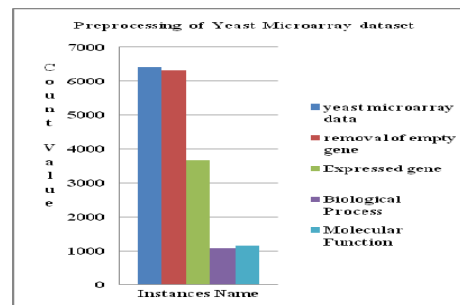


FIG. 2 INSTANCES OF PREPROCESSING

Temporal Associations

The temporal associations of gene for 1153 gene of Molecular function and 1079 gene of Biological Process are extracted using Apriori algorithm for time stamps $\Delta=2$ and $\Delta=4$. The association rule mining algorithm is applied to Biological Process and Molecular Function genes for various support measures 50, 80, 100 respectively. The temporal association rules are extracted for interval 2 and 4 for the Molecular Function and Biological Process for support 50 with 239 and 236 rules and for support 80 with 281 and 355, and for 100 with 261 and 329 rules which is shown in FIG 3.

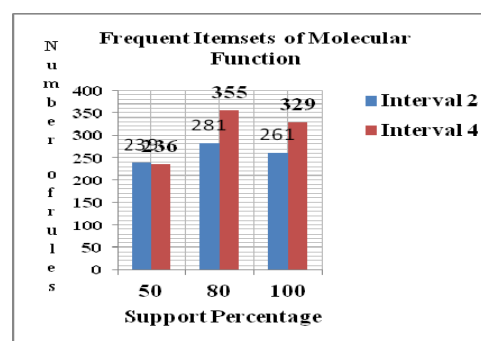
FIG. 3 TEMPORAL RULES OF MOLECULAR FUNCTION FOR $\Delta=2$, $\Delta=4$

TABLE 6 TEMPORAL ASSOCIATION RULES FOR SUPPORT MEASURES

S.no	Molecular Function			Biological Process		
Interval	50%	80%	100%	50%	80%	100%
2	239	281	261	239	281	261
4	236	355	329	236	360	329

The results of the proposed work extracted with the set of rules for Molecular Function and Biological Process

for the time interval 2 and 4 for the co-occurrence of genes are discussed in section 5.

Discussion

The temporal association rules extracted for the Molecular function for the time intervals 2 for the support 50, 80, 100 for the co- occurrence of genes are presented. TABLE 6 provides the snapshot of set of co-occurrence of gene. The gene YALO18CD which is down regulated associated with the genes like YALO27WD, YAL037WD, YALO65CU, YAR009CD, YARO23CU, YARO61CD, YARO62WU; YARO64WD which are up and down regulated is present for support 50%. The same set of gene is found to occur for support 80%. The co-occurrence patterns are not found for support of 100. Likewise gene YAL027WD with down regulation is associated with the genes YARO10CD, YARO10CU, YBL005WD, and YBLOO5WU which are up and down regulated. For support 80%, the gene YAL027WD with down regulation is also associated with the genes YARO10CD, YARO10CU, and YBLOO5WD with up and down regulation. This set of co-occurrence is not found for 100% support. Likewise, the gene YAL037WD with down regulation is associated with the genes YAL065CD, YAR009CU, YAR010CD and YARO64WU with up and down regulation. This association rule is found to occur for support 80% but not in the 100%.

TABLE 7 TEMPORAL RULES OF INTERVAL 2 OF MOLECULAR FUNCTION

S. no	Temporal Transaction Set for 60%	Temporal Transaction Set for 80%	Temporal Transaction Set for 100%
1	YALO18CD--> YAL027WD,YAL037WD, YALO65CU,YAR009CD, YAR023CU,YAR061WD, YAR062WU,YAR064WD.	YALO18CD--> YAL027WD,YAL037WD, YALO65CU,YAR009CD, YAR023CU,YAR061WD, YAR062WU,YAR064WD.	YALO18CD--> YAR010CD,YAR010CU, YAR029WD,YAR029WU, YBL005WU,YBL010CD, YBL010CU,YBL029WD
2	YAL027WD--> YAR010CD,YAR010CU, YBL005WD,YBL005WU.	YAL027WD--> YAR010CD,YAR010CU, YBL005WD,YBL005WU, YBL010CD,YBL010CU, YBL029WD,YBL029WU.	YAL027WD--> YAL065CD,YAR009CU, YAR023CD,YAR029WU, YBL081WU,YBL086CU, YBL111CD,YBL112CU
3	YAL037WD--> YAL065CD,YAR009CU, YAR010CD,YAR064WU.	YAL037WD--> YAL065CD,YAR009CU, YAR010CD,YAR064WU.	YAL037WD--> YAR023CD,YAR062WD, YBL081WU,YBL104CU.
4	YAL064CD--> YAR064WD,YAR064WU, YBL066CD,YBL066CU, YBL095WD	YAL064CD--> YAR064WD,YAR064WU, YBL066CD,YBL066CU	YAL064CD--> YAL065CD,YAR009CU, YAR010CD,YAR029WD

The temporal association rules extracted for the Molecular Function for the time intervals 4 for the

support 50, 80,100 for the co- occurrence of genes are presented. The TABLE 8 provides the snapshot of set of co-occurrence of gene. The gene YARO29WD which is down regulated associated with the genes like YARO62W(DUUU) and YBLO95W(DUDU) and YBRO28C (DUUU) with up and down regulation is found to occur in the 50%. The same set of gene is found to occur for the support 80% and 100%. Similarly, the second set of gene for the support 50 % gene YBR138CD which is down regulated is associated with the genes YBR285W(DUUD), YDL176W(DDUU) with up and down regulation. The same set of gene is found to occur for the support 80% and 100%.

TABLE 8 TEMPORAL RULES OF INTERVAL 4 OF MOLECULAR FUNCTION

S. no	Temporal Transaction set for 50%	Temporal Transaction set for 80%	Temporal Transaction set for 100%
1	YARO29WD→YARO62WD, YARO62WU, YARO62WU, YARO62WU, YBL095WD, YBL095WU, YBL095WD, YBL095WU, YBR028CD, YBR028CU, YBR028CU, YBR028CU	YARO29WD→YARO62WD, YARO62WU, YARO62WU, YARO62WU, YBL095WD, YBL095WU, YBL095WD, YBL095WU, YBR028CD, YBR028CU, YBR028CU, YBR028CU	YARO29WD→YARO62WD, YARO62WU, YARO62WU, YARO62WU, YBL095WD, YBL095WU, YBL095WD, YBL095WU, YBR028CD, YBR028CU, YBR028CU, YBR028CU
2	YBR138CD→YBR285WD, YBR285WU, YBR285WU, YBR285WD, YDL176WD, YDL176WD, YDL176WU, YDL176WU	YBR138CD→YBR285WD, YBR285WU, YBR285WU, YBR285WD, YDL176WD, YDL176WD, YDL176WU, YDL176WU	YBR138CD→YBR285WD, YBR285WU, YBR285WU, YBR285WD, YDL176WD, YDL176WD, YDL176WU, YDL176WU
3	YDL124WU→YDL176WD, YDL176WD, YDL176WU, YDL176WU, YDL180WD, YDL180WU, YDL180WD, YDL180WU	YDL124WU→YDL176WD, YDL176WD, YDL176WU, YDL176WU, YDL180WD, YDL180WU, YDL180WD, YDL180WU	YDL124WU→YDL176WD, YDL176WD, YDL176WU, YDL176WU, YDL180WD, YDL180WU, YDL180WD, YDL180WU
4	YBR147WD→YBR194WD, YBR194WU, YBR194WD, YBR194WU, YCL002CD, YCL002CU, YCL002CU, YCL002CU	YBR147WD→YBR194WD, YBR194WU, YBR194WD, YBR194WU, YCL002CD, YCL002CU, YCL002CU, YCL002CU	YBR147WD→YBR194WD, YBR194WU, YBR194WD, YBR194WU, YCL002CD, YCL002CU, YCL002CU, YCL002CU

The data set consists of 1153 gene of Molecular Function and 1079 gene of Biological Process. It is seen that only 1132 genes of Molecular Function gene are associated with other genes in the corresponding support measures and 21 genes are not found to co-occur with other genes like YGL057C , YGL081W, YPL030W, YCR051W, YDR520C, YDR222W, YDR266C, YJL149W, YPL183C, YLL017W, YFR006W, YPR152C, YIL177C, YJL149W, YNR021W, YHR177W, YKR011C, YDR287W, YKL033W, YDR415C, YDR539W.

On the other hand, the Biological Process of 1079 gene, a count of 1061 genes co-occur with other genes in its corresponding support measures and the remaining 18 genes are not found to occur with other genes like YGL057C, YGL081W, YPL030W, YDR266C, YJL149W, YPL183C, YLL017W, YFR006W, YPR152C, YIL177C, YJL149W, YNR021W, YHR177W, YKR011C, YDR287W,

YKL033W, YDR415C, YDR539W.

The temporal association rules extracted for the Molecular function and Biological Process of time interval 2 and 4 are compared to analyze the similar pattern of rules. The temporal association rules extracted for the time intervals 2 and 4 for the Molecular Function and Biological Process genes using the Apriori algorithm are taken and the similar patterns of rules for the time intervals 2 and 4 for the functionalities are analyzed.

The similar pattern of rules extracted for the support 50% with interval 2 is 153 rules, support 80% is 154 and 142 rules for the support 100%. Likewise, the rules extracted for the interval 4 for support 50% is 53, support 80% is 132 and 100% is 163 rules. The similar pattern of association rules for the Molecular function consisting of 1153 genes and 1079 genes of Biological Process extracted is shown in the FIG 4.

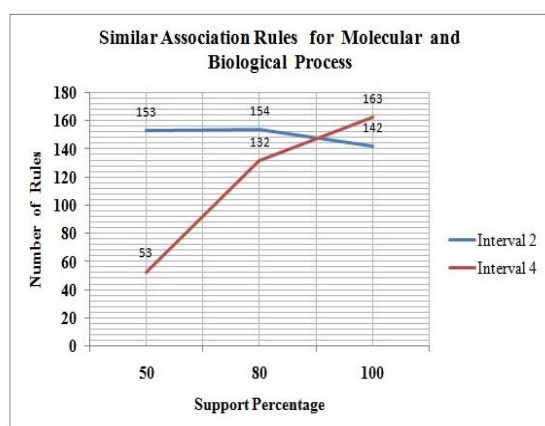


FIG. 4 SIMILAR PATTERN OF RULES OF TIME INTERVAL 2

The similar pattern of rules extracted for the Molecular function and Biological Process for the time intervals 2 and 4 is shown above. It is seen that the rules extracted for time interval 2 seems to be increasing as 153 rules for the support 50% and 154 for the support 80%.

A slight variation in the rules which get decreased for the support 100% with 142 rules. In case of time interval 4, the similar pattern of rules starts with 53 rules for the support 50, a sudden increase in the similar pattern of rules for the support 80 with 132 rules and 163 rules for the support 100.

Conclusion

Microarray technology is a tool which helps us to analyze the expression levels of thousands of genes simultaneously. To analyze the gene expression data, association rule mining is used to discover the frequent patterns, co-relations of the microarray genes.

The proposed method uses the Yeast *Saccharomyces cerevisiae* dataset to find the association using the Apriori algorithm in different experimental conditions for time interval 2 and 4. The original microarray dataset consists of two functionalities called the Molecular function with 1153 gene and Biological Process with 1079 gene. The dataset is processed to extract the temporal association rules for the time interval 2 and 4 for the support measures. The temporal association rules extracted for the Molecular function and Biological Process for the time interval 2 for the support 50% is 239 rules, 80% is 281 rule and for support 100% is 261 rules. Likewise, the rules extracted for the time interval 4 for the support 50% is 236 rules, 80% is 360 rules ,100 is 329 rules. A slight difference occurs in the support 80% in the interval 4 of Biological Process and Molecular Function. When the extracted temporal association rules are compared for the Molecular function and Biological Process, it is found that 80% of temporal rules are same.

The extracted temporal association rules for the Molecular function and Biological Process are taken and analyzed to extract the similar pattern of rules. The similar pattern of rules extracted for the support 50, 80,100 for the time interval 2 seems to be decreased as 153,154,142 rules. In case of time interval 4, the rules seems to be increased as 53, 132,163 rules.

The biological inference of similar pattern of rules for the time interval 2 seems to be decreased and increased in interval 4. The similar pattern of rules extracted for the support measures is again analyzed to find the non associated gene. It is found that a number of 21 gene are found to be irrelevant to other genes in the Molecular function and in the Biological Process, as well as a number of 18 gene The non-association genes for the Molecular function and Biological Process are YGL057C, YGL081W, YPL030W, YDR266C, YJL149W, YPL183C, YLL017W, YFR006W, YPR152C, YIL177C, YJL149W, YNR021W, YHR177W, YKR011C, YDR287W, YKL033W, YDR415C, YDR539W. When the temporal rules of the proposed work are in analysis, it is found that 80% of the similar patterns of temporal association rules of the Molecular function and the Biological Process for the intervals 2 and 4 are same. The biological validation is done in the Future work.

REFERENCES

Agrawal R., Imielinski T., Swami A., 1993, "Mining Association Rules between Set of Items in Large Databases",

Proceedings: ACM SIGMOD International Conference on Management of Data.

Anandhavalli Gauthaman. M, 2008, "Analysis of DNA Microarray Data using Association Rules: A Selective Study", Cited from: <http://citeseerx.ist.psu.edu/viewdoc/>.

Anour F.A., Dafa-Alla, Ho Sun Shon, 2010, "IMTAR: Incremental Mining of General Temporal Association Rules", Proceedings: Journal of Information Processing Systems.

Arun.K.Pujari, 2001, "Data Mining Techniques", Universities Press (India) Limited, ISBN-81-7371- 3804.

Attila Gyenesei, Ulrich Wagner, Simon Barkow-Oesterreicher, Etzard Stolte and Ralph Schlapbach, 2007, "Mining Co-Regulated Gene Profiles for the Detection of Functional Associations in Gene Expression Data" Proceedings: Oxford Journals.

Cheung D.W., 2010, "Maintenance of Association Rules in Large Databases: An Incremental Updated Technique", Cited from: <http://scholar.google.co.in/>.

Christian Borgelt, 2003, "Efficient Implementations of Apriori and Éclat", Cited from: <http://scholar.google.co.in>.

Elizabeth Georgii, Lothar Richter, 2005, "Analyzing Microarray Data Using Quantitative Association Rules", Cited from: <http://bioinformatics.oxfordjournals.org>.

E.papageorgiou. I. Kotsioni, A. Linos, 2005, "Data Mining: A New Technique in Medical Research", Cited from: <http://www.enggjournals.com/>.

Gaurav Pandey, Gowtham Atluri, Vipin Kumar, 2008, "Association Analysis for Real-Valued Data: Definitions and Application to Microarray Data", Cited from: <http://www-sers.cs.umn.edu/>.

Hojung Nam, Kiyoun Lee and Doheon Lee, 2009, "Identification of Temporal Association Rules from Time-Series Microarray Data Sets", Cited from: <http://www.biomedcentral.com>.

Junheng Huang, Wang Wei, 2007, "Efficient Algorithm for Mining Temporal Association Rule", Published in International Journal of Computer Science and Network

Security. (IJCSNS), Cited from: http://paper.ijcsns.org/07_book/200704/20070438pdf.

Manjula Kurella, Li-Li Hsiao, Takumi Yoshida, Jeffrey D. Randall, Gary Chow, Satinder S. Sarang, Roderick V. Jensen and Steven R. Gullans, 2001, "DNA Microarray Analysis of Complex Biologic Processes", Vol: 12, 1072–1078, Cited from: <http://jasn.asnjournals.org/content/12/5/1072>.

Pavel Smrř and Jana řilhavá, 2008, "A Framework for Microarray Data Mining and Processing", Cited from: <http://citeseerx.ist.psu.edu/>.

Xiang-Rong Jiang, Le Gruenwald, 2005, "Microarray Gene Expression Data Association Rules Mining Based on BSC-Tree and FIS-Tree", Cited from: <http://dmlab.mgt.ncu.edu.tw>.

Xiang-Rong Jiang, Le Gruenwald, 2005, "Microarray Gene Expression Data Association Rules Mining Based on BSC-Tree and FIS-tree", Cited from: <http://dmlab.mgt.ncu.edu.tw>.

Zhai Liang, TANG Xinming, LI Lina, J IANG Wenlianga, 2006, "Temporal Association Rule Mining Based on T-Apriori Algorithm and Its Typical Application", Cited from: <http://isprs-wgii-1.casm.ac.cn/source/Algorithm.pdf>.

Bhuvaneswari V. completed Ph.D and M.Phil in Computer Science at Bharathiar University, Masters Degree (MCA) in Computer Applications in IGNOU and the Bachelors Degree (B.Sc.) in Computer Technology PSG Technology. She has qualified UGC-JRF in the year 2003. Currently she has been working as Assistant Professor in department of Computer Application, Bharathiar University, Coimbatore, India since 2005. Her research interest includes Data Mining, Computational Biology, Bioinformatics and Evolutionary Computing. She has authored more than 30 papers in Journals and Conference and guided 15 research scholars for M.Phil in computer science and currently guides research scholars at Bharathiar University in the area of Data mining and Bioinformatics.

Umajothy P. completed M.Phil in Computer Science at Bharathiar University, M.sc (Computer science) in Sree Saraswathi Thyagaraja College and the Bachelor degree (B.Sc) in Computer science in the Sree Saraswathi Thyagaraja College, pollachi. Research Interest includes in the area of data mining and bioinformatics.